Y2 ANALYTICS

MEMORANDUM

To: Interested individuals
From: Scott Riding, Kelly Patterson, and Quin Monson, Y2 Analytics
Date: October 28, 2014
Re: Survey sample size and its effect on accuracy

Sample size – or the number of people who participate in a survey – is not the primary determinant of survey accuracy. This is a common misconception. Occasionally there are commentators who criticize public polling because it only includes a few hundred interviews to make its estimates for a population of thousands or millions.

But a sole focus on the number of interviews is misplaced because it misses an even more critical component – do the interviewees represent a random sample of the population as a whole? Sample size means very little if the individuals included in the sample are not demographically and geographically representative. This makes sense intuitively. Consider a fictional South Jordan City citizen survey of 5,000 interviews in which the respondents consist solely of citizens who are home during the day, or only those who regularly read a local publication. Few would claim that this group of interviews represents the city because it would be systematically biased toward one demographic group or another.

Despite its error, this approach to surveying is common. A timely example is the Jordan School District research on the 2012 school bond. In June, the school district published results of nearly 8,000 interviews and asserted that the bond had 80% support and would easily gain electoral approval. However, these interviews were not collected as part of a random sample. Instead, the survey was made available to every household in the district and its existence was especially communicated to parents and educators. As a result, families and professionals with a stake in the outcome participated at disproportionate levels in the survey. Unfortunately, that group was not a representative sample of the voters and in November the bond failed by a large margin.

We employ scientific random sampling to ensure our respondents represent the entire city. Not everyone will be selected, but we ensure that those who are selected are given a convenient and accessible way to participate. We follow up with selected individuals who do not respond to our first invitation. Our experience in a variety of municipal contexts gives us confidence that this mode does a good job of obtaining a random sample – favoring no demographic or geographic group – at the lowest possible cost to the city budget.

Once you have a random sample, the size of that sample does affect the precision of the results. The margin of error or confidence interval of a probability survey is mostly a function of sample size. But the relationship is not 1:1. In other words, investing more resources to increase the sample size (and decrease the margin of sampling error) is a good use of resources but this investment has diminishing returns that are illustrated by the figure below. A sample size of 500 has a margin of error of about 4.4%. When you double the sample size from 500 to 1000, the margin of error goes down from 4.4% to 3.1%. This is a significant reduction to be sure, but doubling the sample size also roughly doubles the data collection costs.

The proposed sample size simply reflects a judgment about the likely analysis of interest to South Jordan City, the anticipated length of the survey questionnaire, and the resources available to conduct the survey. We believe that 500 interviews is the best tradeoff between costs and benefits for South Jordan.

MARGIN OF ERROR AND SAMPLE SIZE